Climate change since 1750

Project Final Paper

Kaixin Chen
Department of Computer Science
University of Colorado Boulder
1111 Engineering Dr
Boulder, Colorado 80303
changshawuyanzu@gmail.com

Xinyi Zhang
Department of Computer Science
University of Colorado Boulder
1111 Engineering Dr
Boulder, Colorado 80303
xizh3366@colorado.edu

Yonghao Zhang
Department of Computer Science
University of Colorado Boulder
1111 Engineering Dr
Boulder, Colorado 80303
yozh2004@colorado.edu

ABSTRACT

Climate change, other than global warming, is a significant change of weather pattern that will last for an extended period of time. It includes global warming, global cooling, extreme weather on the certain area, and many other unusual weather changes. Human influences are considered as the dominating factor of climate change. The human-induced changes are so powerful that they exceed the range of natural variability. The purpose of this project is to demonstrate the relationship between climate change and regional issue and to prove whether global warming is a fact or friction. Through analysis of the earth surface temperature data through 1750 to 2016, it is possible to find that the climate change is a fact and the regional issue because of human-induced changes does influence the climate.

1 INTRODUCTION

In this paper, we will examine the trend of climate change since 1750 in different area and country all over the world. We map the climate data with average temperature, country, latitude, and longitude while also concerning the uncertainties of each numerical data. We aim to answer the following questions in the course of the project:

(1) Is Global Warming a fact or fiction?
(2) Is temperature rising a global phenomenon or it only happens in the certain area?
(3) What is the relationship between the climate change and altitude?
(4) What is the relationship between the seasonality and climate trend?
(5) What is the relationship between the highs and lows trend and climate trend?

To solve these questions, we will explore the relationship between temperature and latitude and discover the land average high and low trends of the climate. We would also concern the effect of melting glaciers and the development of industry to have a more conclusive ideal about the climate change. The results from this analysis could help those climate studies discovering the phenomenon of global warming and helping those people that are not familiar with the climate change get a better and straightforward understanding about the changes in climate these years.

2 RELATED WORK

There are many existing studies analyzing how global climate has changed, which includes analysis on temperature, human effects, and some natural factors. These studies also concluded how to solve or slow down the process of the climate change and what consequences the climate change might have.

2.1 EPA: United States Environmental Protection Agency

The report of EPA claimed that the primary cause of climate change is imbalance between energy entering and leaving the planet's system. As for we know, when incoming energy from the sun is absorbed by the Earth system, Earth warms and vice versa. The imbalance of energy is caused by the greenhouse effect, natural changes including changes within the sun and changes in Earth's orbit and changes in reflectivity. Greenhouse effect that causes the atmosphere to retain heat. Natural changes affect can affect the intensity of the sunlight that reaches Earth's surface that affects how much energy reaches Earth's system. Finally, changes in reflectivity also affect how much energy enters Earth's system because the amount of sunlight that reaches Earth can be reflected or absorbed that depends on Earth's surface and atmosphere. [1]

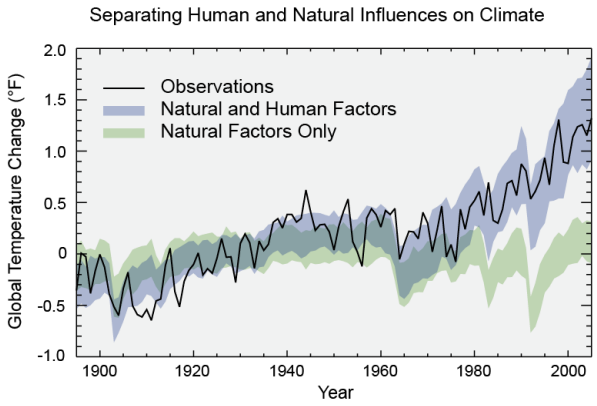

Figure 1: Human and Natural Influence on Climate

2.2 NASA: Climate Change and Global Warming

The NASA research about the climate change also concludes the global warming. The research conclusion of NASA could be regarded an authentic Wikipedia; in other words, the research includes evidence, causes, effects, and solutions. We might focus

our attention on the Climate Resource Center, which including the analysis of the climate change through many aspects. These results are useful for us to prove the credibility of our project and could also help us conclude many other aspects we might need in the final conclusion part. [2]

3 DATA SET

We obtained the five global temperatures since 1750 datas from the Kaggle.com. [3] The raw data comes from the Berkeley Earth data page, which is affiliated with Lawrence Berkeley National Laboratory. These five datasets repackaged 1.6 billion temperature reports from 16 pre-existing archives.One of the most noticeable feature of this dataset is the temperature data from 1750 to 1850 has great uncertainties. Cleaning process is extremely important for our data analysis process in this case. In this dataset, we have include several files:

- Date: starts in 1750 for average land temperature and 1850 for max and min land temperatures and global ocean and land temperatures
- LandAverageTemperature: global average land temperature in celsius
- LandAverageTemperatureUncertainty: the 95% confidence interval around the average
- LandMaxTemperature: global average maximum land temperature in celsius
- LandMaxTemperatureUncertainty: the 95% confidence interval around the maximum land temperature
- LandMinTemperature: global average minimum land temperature in celsius
- LandMinTemperatureUncertainty: the 95% confidence interval around the minimum land temperature
- LandAndOceanAverageTemperature: global average land and ocean temperature in celsius
- LandAndOceanAverageTemperatureUncertainty: the 95% confidence interval around the global average land and ocean temperature

Other files include:

- Global Average Land Temperature by Country (GlobalLandTemperaturesByCountry.csv)
- Global Average Land Temperature by State (GlobalLandTemperaturesByState.csv)
- Global Land Temperatures By Major City (GlobalLandTemperaturesByMajorCity.csv)
- Global Land Temperatures By City (GlobalLandTemperaturesByCity.csv)

4 MAIN TECHNIQUES APPLIED

4.1 Preprocessing & Data Cleaning

For the data process, at the beginning, we found the raw data we need from Kaggle and we separated some of the big datasets in small parts to make it acceptable by Excel. After we open it on Excel, we noticed there are missing values and possibly duplicate data in our raw datasets. Then we defined the type of attributes for each raw data set by Excel and imported the raw data correctly in MySQL database which is the main tool we use for data cleaning and data warehousing. The way we handle with the missing values is we set the default value of variables in our table stored the raw data to NULL which avoid the possible calculation errors caused by missing values. For duplicate data, we used the count function provided in MySQL to find them out and delete them from our dataset tables. Simultaneously, since we stored them in the separate table first, we did the data integration by join function then see if there is duplicated data/primary key related to two raw data sets in MySQL also. As the result of data integration, we have a huge table stored the whole dataset and the running time of our source code on it is kind of unacceptable. We did data reduction by select certain type of data which we needed for code running in our big dataset and create a new table for them. The benefit with data reduction is the running time of our code is significantly improved and we could visualize them in Excel easily. Obviously, the dataset imported from raw data does not meet our needs. For example, the temperature scale in our raw data is not the same. We have to do Fahrenheit to Celsius conversion for some temperature attributes. Besides, the latitude attributes were stored in varchar type which includes alphabet N/S. It required us to store the number and alphabet in the separate variable when we trying to do further research on the relationship between temperature rising and latitude which required us to do data transformation. We also did some data aggregation to store the annual temperature or the average temperature in a certain period.

4.2 Data Warehouse

For Data Warehousing, the main tool we used is MySQL because itfis high performance and on-demand scalability. Since MySQL does not support image generation and our group is measuring the image generation on Matlab, we transferred and extracted the dataset in the acceptable format in Matlab to help us do further works. Data imported from original site is stored in separated table and we have new tables which are results of data integration and data reduction. All the tables in our dataset were well sorted and cleaned.

4.3 Classification & Clustering

For classification, We used the WEKA which is a data mining software in java. It not only has powerful function which includes various algorithms for classification and clustering but also allows us to connect WEKA with our MySQL database which made the data processing easily. To be more specific, we used the Naive Bayes function for classification and SimpleKmean function with Euclidean distance function as the distance function for clustering. The result of classification and clustering helped us to build the scale on latitude for the further research.

5 TOOLS

- **SQL**
  Helping host the preprocessed database.
- **Matlab**
  The tool we used at the beginning to analysis the data and

plot the graph. It was also a great tool to do the fitting process.

- **Weka**
  Containing tools for data pre-processing, classification, regression, clustering, association rules, and visualization.
- **R**
  Use R to visualize our final result. It has great packages that could form graphs neatly.
- **Excel**
  The application that stores the original data.
- **Latex**
  Using LaTeXto form the report.

6 RESULT

6.1 Global Temperature Analysis

For this part, we analysis the dataset fiGlobal Land and Ocean and Land Temperaturesfi in order to obtain the global average temperature from 1750 to 2015. In Figure 2, all these three lines have great vibration at the left side of the graph. This fact was caused by the great uncertainties before 1850. In this case, even though the peak value of the global maximum temperature occurs around 1765, we still put more attentions on the values after 1850. Both global minimum temperature and global average temperature have a tendency of increasing. The global maximum temperature is relatively steady; however, starting from 1980, the maximum temperature increases suddenly.

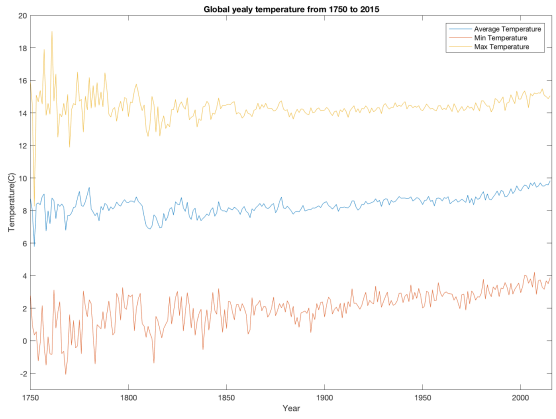

Figure 2: Global Land Temperatures

In Figure 3, we tried to apply the method of modelization for the global yearly average temperature. The polynomial function indicates that the global average temperature decreases a little bit around 1840. Both the linear function and polynomial function could tell the global yearly average temperature keeps increasing. The above graphs include only land temperature. To analysis the global phenomenon, it is necessary for us to consider the factor of the ocean.

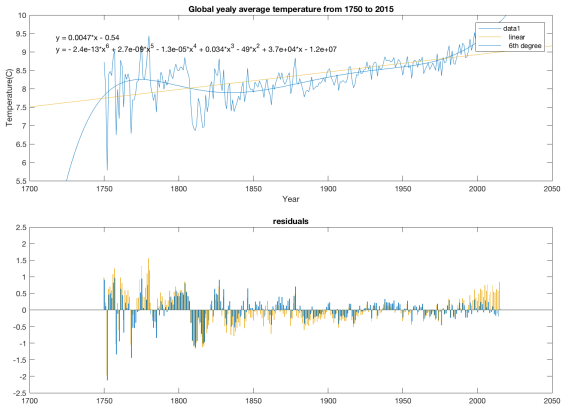

Figure 3: Global Land Temperatures with Fitting

Figure 4 tells the land and ocean average temperature from 1848 to 2015. The date starts from 1848 as the dataset does not include most of the data before 1848. The land and ocean average temperature is much higher than the land average temperature. The main factor that caused this difference is the heat capacity of water. The ocean is much more stable than the land and also takes a longer time to cool or warm. As the fitting model implies, the land and ocean average temperature has a tendency of increasing too.

Conclude what we obtain above, it is not hard for us to find that

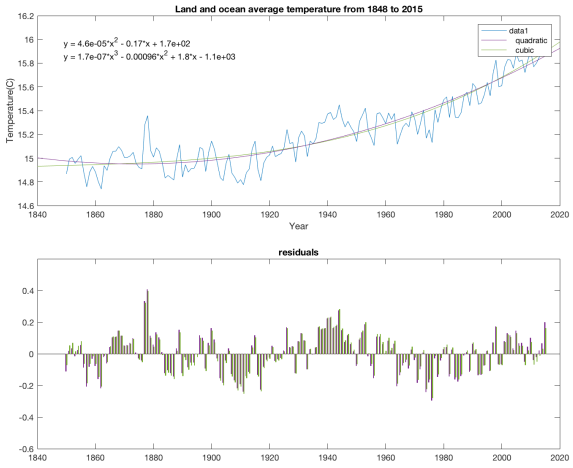

Figure 4: Global Land and Ocean Average Temperatures with Fitting

the global average temperature, including the land temperature and ocean temperature, is increased during recent 200 years. This is a great evidence that proves the global warming but we could not conclude that the climate is warming only according to the global average temperature.

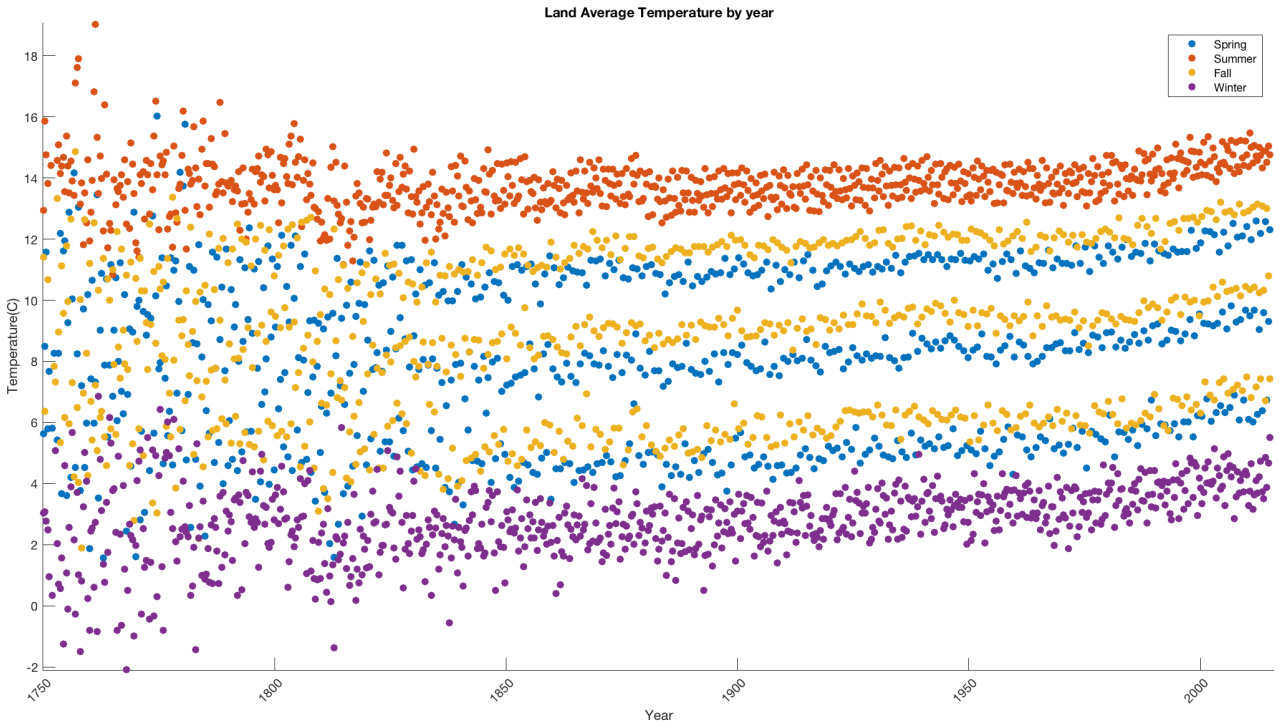

Figure 5: Land Yearly Average Temperature by Season

6.2 Seasonality

In Figure 5, different colors represent different seasons. Summer is placed at the top of the graph, which also representing the highest temperature range. Some dots seem to be errors; however, like the same reason of uncertain data before 1850, these dots are influenced by the uncertainty actually. The right side of the scatter plot was much denser than the left side, which indicates that those data is more precise and temperature during these years is changed little by little. The most significant information we could learn from this graph is the temperature keeps rising in all four seasons. And the rising rate was similar during the recent 200 years. This graph could help us answer the question "*What is the relationship between the seasonality and climate trend?*" The global temperature does have a strong dependence on seasonality, like the summer always has the hottest temperature and winter has the coolest temperature. Nevertheless, the temperature variables of these four seasons show a trend towards warmer weather in different degree.

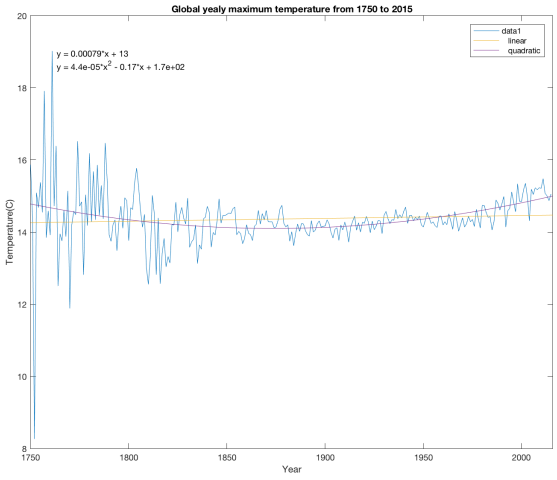

Figure 6: Global Yearly Maximum Temperature

6.3 High Trends & Low Trends

As we could see in Figure 6, The maximum average temperature is relatively stable after 1850. The linear model looks like a straight line but is actually a rising line according to the equation

$$y = 0.00079x + 13 \tag{1}$$

The quadratic model seems to be a gradual parabola because of those uncertain values before 1850. The period between 1830 and 1930 is the low trends of the high temperature. Despite the left side of the graph, we can find the slope on the right side becomes steeper and steeper as time goes. This means the rising rate of the high trends is increasing these years.

The minimum average temperature in Figure 7 is much more significant in some case. Even though the vibration is great especially on the left side of the graph, we could still find the high trends

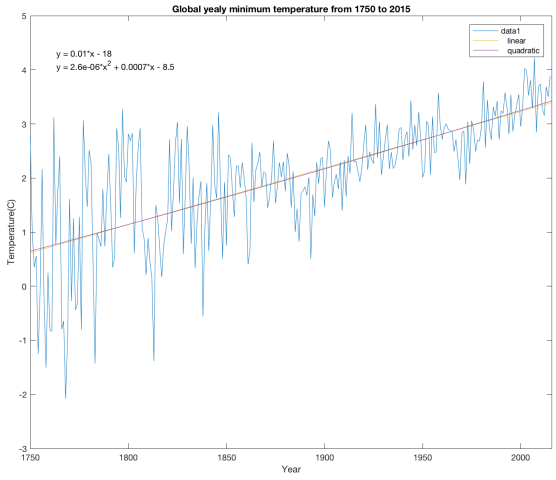

Figure 7: Global Yearly Minimum Temperature

through modelization. The linear model and quadratic model appear to overlap but when we zoom in the graph, we find that the line of the quadratic model is above the line of the linear model, which tells the rising rate of low temperature is increasing in recent 50 years.

6.4 Regional Analysis

The Histogram of Major City Average Temperature in Figure 8 is the analysis of Average temperature of over one hundred cities from 1849 to 2013. Basically, we summarized the yearly average temperature of each major city from 1839 to 2013, which is about two hundred years, and then we used **Matlab** to calculate the average of their summation. The method we used is probably not very accurate because summarizing average temperature of major cities can just show the average temperature of different countries or even different continents. But we can use these outcomes to calculate the standard deviation of the average temperature of each major city or country or continents and then conclude which city or which country or which continent get the greatest variance. Besides, we can compare latitudes and longitudes of each major city to see how location influence the average temperature and we will try to figure out some specific coefficients. Finally, we will use this dataset to calculate the average growth rate of the temperature of each major city by regressing the data to a linear function.
We all know the longitude and latitude of each major city inevitable associate the average temperature of each city. But there are something we usually ignore. For instance, the emission of greenhouse gas can not only damage the Ozone layer but also influence the global precipitation.

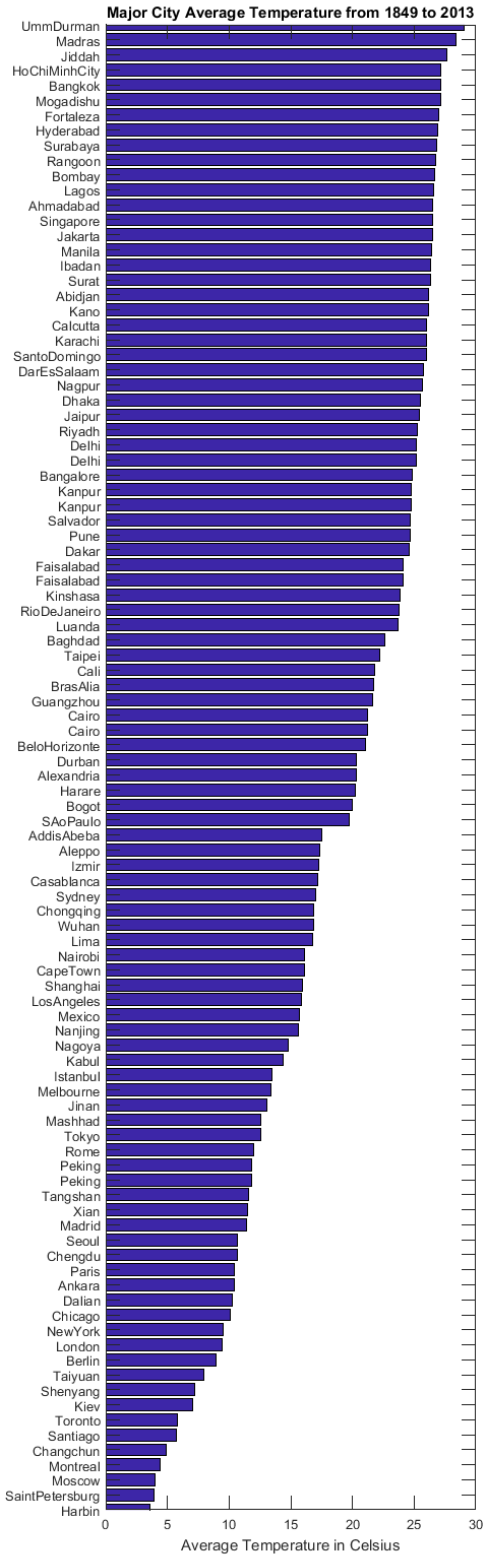

Figure 8: Major City Average Temperature from 1849 to 2013

The damage of the good Ozone layer will reduce the protection of lives on Earth from the sun's harmful ultraviolet (UV) rays and will increase the absorb of heat from the sun. On the other hand, the location of precipitation changes also influence the global surface temperature. For example, if the precipitation can't reach on the land, then the heat absorb from the sun could not be dissipate by evaporation. Therefore, these two aspects will be treated as the most important goal for as to find out in the next few weeks.
As the human-induced changes is the dominating factor that cause climate change, we also compared the temperature change of a developing country and a developed country. China is the largest developing country with a large population. We compared the temperature in 1949, the period after war and before extensive industrial development in China, with current temperature as Figure 9,10, and 11 shown.

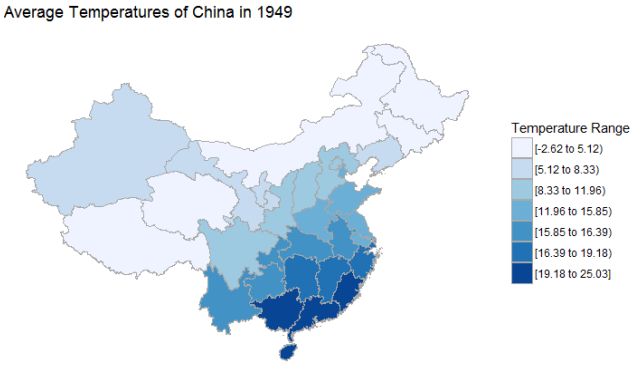

Figure 10: Average Temperature of China in 1949

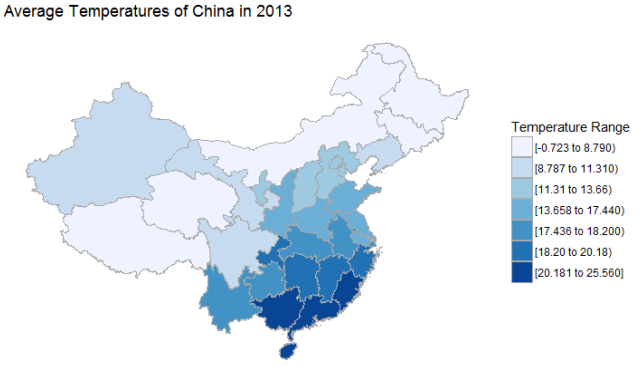

Figure 11: Average Temperature of China in 2013

As we could observe from Figure 11, the upper part of China changed most during these years and the coastal part was relatively stable.We zoomed in each area and chose several representative area to do the further analysis.

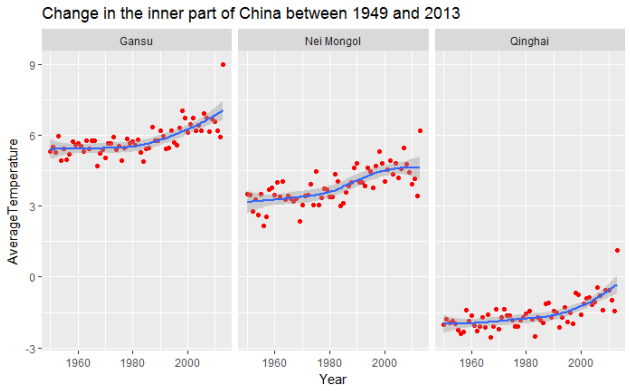

Figure 12: Change in inner part of China from 1949 to 2013

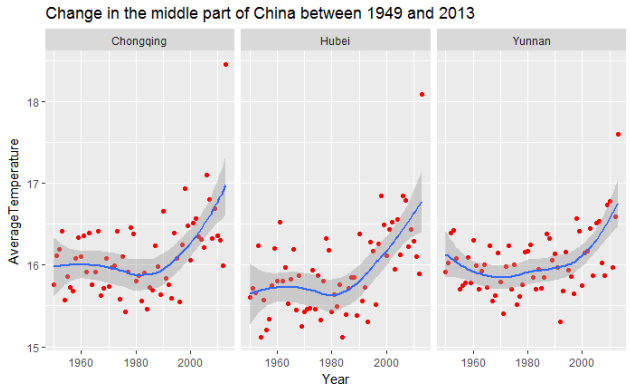

Figure 13: Change in middle part of China from 1949 to 2013

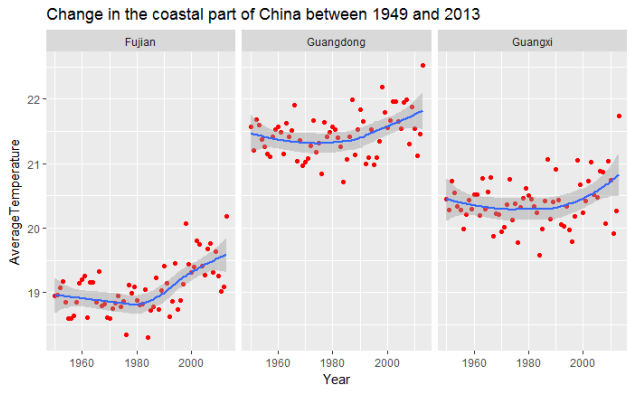

Figure 14: Change in coastal part of China from 1949 to 2013

According to Figure 12, 13, and 14, we found that the rising rate of the temperature in continental parts of China is much greater

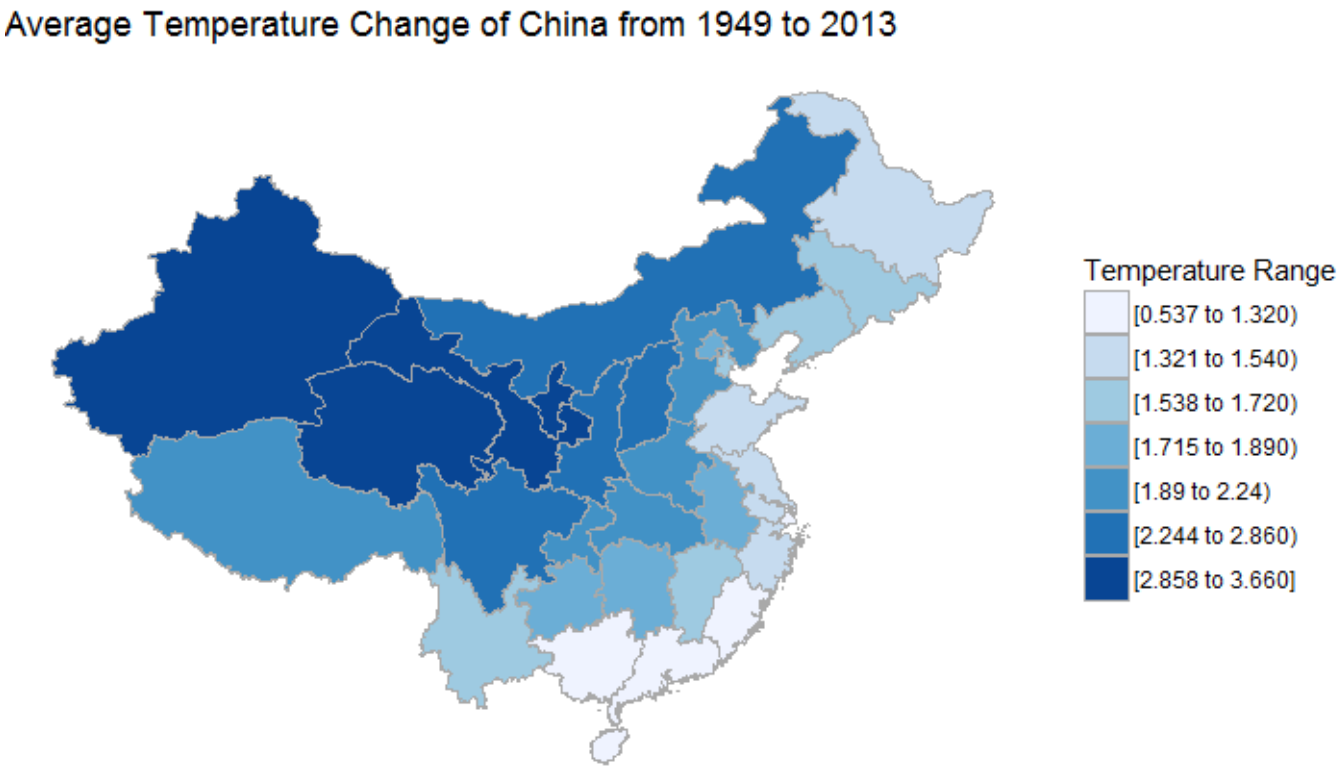

Figure 9: Average Temperature in China from 1949 to 2013

than the rising rate of the temperature in coastal parts. The heat capacity of the water might cause this interesting fact. It takes much longer time for ocean to get heated or cool down; therefore, the coastal area showed a much slower rate of change. Another typical fact we should notice is the continental parts of China are under great development these years. We can find in Figure 11, those newly developing provinces had greatest change of temperature. The temperatures increased exponentially especially around recent years. In this case, we could say the human-induced change does influence the climate change.

Another example we chose was the United States. As the largest developed countries in the world with relatively low population, the United States was a great example for us to analysis the regional effect.

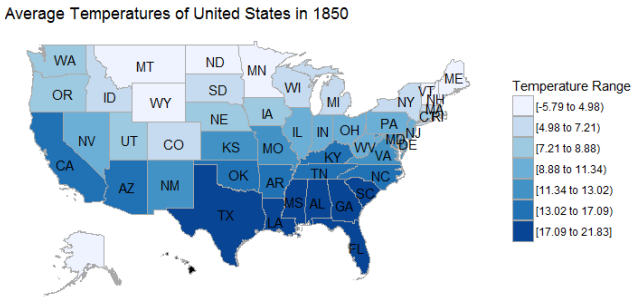

Figure 16: Average Temperature of United States in 1850

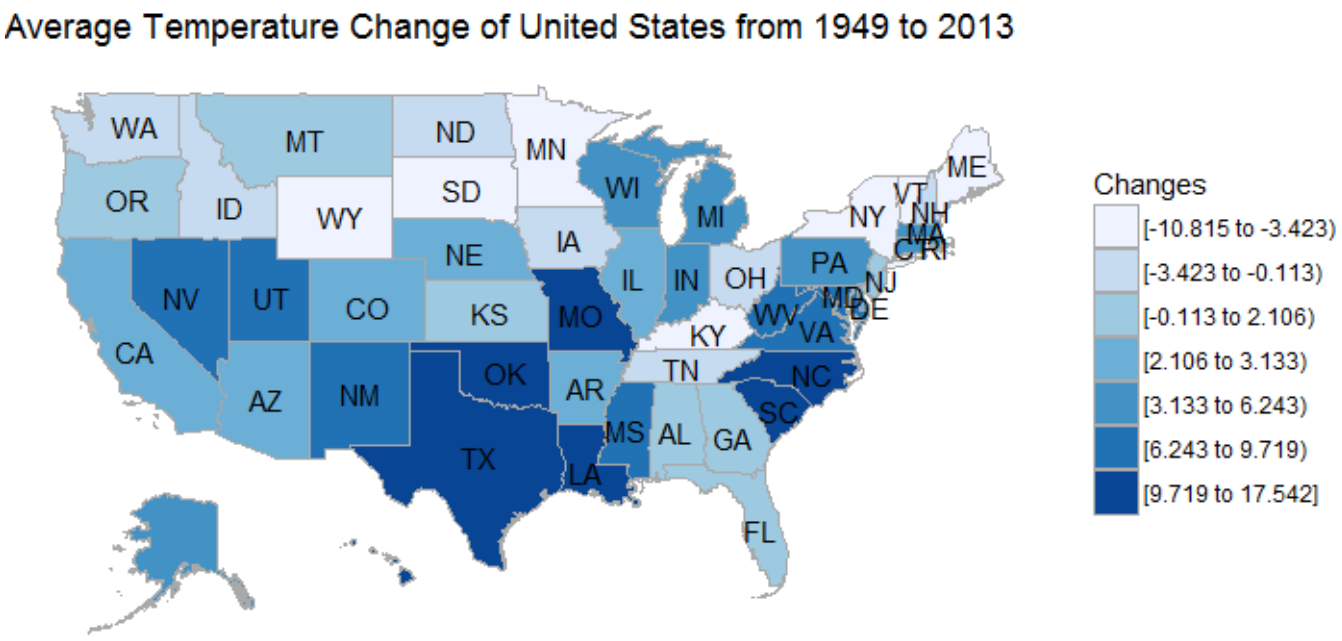

Figure 15: Average Temperature Change of United States from 1850 to 2013

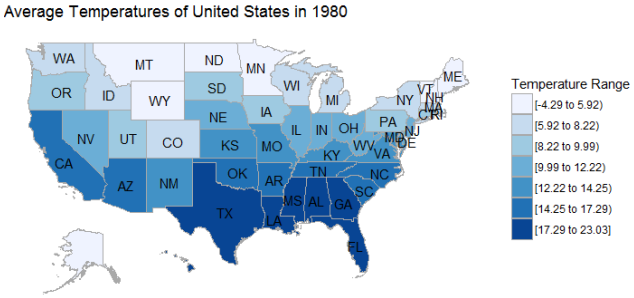

Figure 17: Average Temperature of United States in 1980

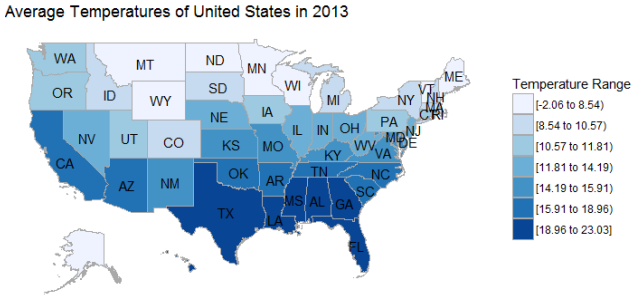

Figure 18: Average Temperature of United States in 2013

We can find the temperature changed area in the United States focused on the southern part and the east coast.The low temperature also increased from -5.79 Celsius in 1850 to -2.06 Celsius in 2013. The change in the United States was kind of smooth but we could still find that the industrial development and city development. In conclusion, the climate change in certain area is also influenced by the regional issue. The developing area and the densely inhabited district have the greatest change because of many human-induced changes, such as industrial pollution and exhaust emission. Also, we should include the urban heat island effect as this effect might also cause the temperature in the urban area to increase.

6.5 Relationship between Climate Trend and Latitude

We examined the relationship between latitude and climate change by comparing the average temperature from 1750 and from 2000 in different latitude areas, including tropic, subtropic, temperate zone and frigid zone, and then we randomly picked four cities as Table 1 to 4 shown. The scale we made for low/mid/high latitude is the combination of our result of classification/clustering and traditional geographic scales.

Table 1: Major city in tropic

Palu (0.8S)				
Season	Winter	Spring	Summer	Autumn
Average Temperature from 1750	20.5	20.7	20.5	20.8
Average Temperature from 2000	26.4	26.8	26.5	26.3

Table 2: Major city in South subtropic

Kimberley (28.13S)				
Season	Winter	Spring	Summer	Autumn
Average Temperature from 1750	23.6	17.1	10.6	19.0
Average Temperature from 2000	24.2	18.0	11.2	19.4

Table 3: Major city in temperate zone

Changchun (44.2N)				
Season	Winter	Spring	Summer	Autumn
Average Temperature from 1750	-14.3	6.0	22.0	5.9
Average Temperature from 2000	-13.1	7.7	22.7	6.9

Table 4: Major city in frigid zone

Norilsk (69.92N)				
Season	Winter	Spring	Summer	Autumn
Average Temperature from 1750	-29.4	-16.2	8.9	-10.7
Average Temperature from 2000	-28.3	-14.5	10.4	-8.6

Obviously, the average temperature in recent 17 years shows an upward tendency and the global warming has a significant impact on low latitude. The average temperature in low latitude area was increased by 2 Celsius degrees and in city Palu, where near the equator, the average temperature was increased by 6 Celsius degrees every season. Besides, based on the data we have, the climate change trend is entirely moved upward

7 APPLICATIONS

Over the past few years, the public's attention to the warming of the Earth has improved significantly. People are highly concerned about the scientific explanations of the causes of global warming, and at the same time political and economic debates raised. Climate is a major factor affecting resource abundance. Resource abundance is not only the constraint of economic decision-making but also the boundary of scientific thinking, even it is a fundamental factor for a country to be considered to carry out macro-control. Among the external factors that affect the abundance of resources, the climate is undoubtedly one of the most important factors. Thatfis why the climate change analysis become the foundation for scientific development decision-making.

The warming of climate warming has become an international consensus; participation and in the future implementation of the "Kyoto Protocol" agreement to become the rise of China's spontaneous and conscious choice. The response to the global action mechanism for climate change began with the Kyoto Protocol. China's participation in the future implementation of the "Kyoto Protocol" agreement, which is the people's spontaneous attention to the people's livelihood needs, is the rapid economic development of China's spontaneous responsibility, but also China to integrate into the global economic international requirements.

Appropriate and stable climate is a necessary condition for human beings to produce and survive in the earth's environment. However, based on the result of our analysis, we found that nearly 200 years, the global climate under the influence of human activities, there have been some unnatural and abnormal changes, these changes can be collectively referred to as climate change. At present, the international community discussed the issue of climate change, which mainly refers to the global warming generated by increase in greenhouse gasses discharge. Greenhouse gasses especially refers to the gasses generated by human society in the production and living process, including the extensive use of coal, oil, natural gas

and other fossil fuels emissions of carbon dioxide gas, resulting in the Earth's atmospheric carbon dioxide concentration gradually increased, the formation of the so-called "carbon dioxide cover", The sunlight carried the heat can get into the earth but the earth can not be dissipate heat, so that the surface temperature of the earth gradually increased, resulting in the phenomenon of climate warming.

Although the maintenance of Earth's surface temperature, it is inseparable from greenhouse gasses. By checking on Google and Wikipedia, we found that if there is no greenhouse gasses, the global average surface temperature should be -18.0 fflfl, and before the industrialization of a long time the average temperature of the global ground is actually about 15 fflfl. Thus, global warming is clearly a catastrophic phenomenon caused by the high concentration of greenhouse gases significantly beyond the limits of nature. If the concentration of greenhouse gasses in the atmosphere continues to increase, further blocking the long-wave radiation emitted by the Earth into space, in order to maintain the radiation balance, the ground will be warming to increase the long-wave radiation. After the increase in ground temperature, water evaporation will increase, the snow will melt, but also to further warming the surface, that is, the formation of positive feedback to make global warming more significant.

The Impact of climate change have been studied mainly in agriculture, water resources, natural environment, and coastline. Besides, the objects of these studies constitute other original elements of economic development. The preliminary results show that climate change will exacerbate the water shortage in many developing countries. Climate change increasingly caused the instability of agricultural production, local high-temperature drought damage and the risk of spring frost. The observation shows that climate warming has led to the interruption of the northern rivers, the advancement of the eastern phenology, the northern boundaries of subtropical and temperate zone moved to Arctic. The influence of the countries in subtropical and temperate zone is greater than the other latitudes on the earth because making heating in winter and make cooling in summer will spend more energy. This problem is globally serious. In addition, in the special ecosystem, due to climate warming and human activities decrease or disappear, the risk of desertification will increase and the forest ecosystem structure will change.

Climate protection involved a significant reduction in greenhouse gas emissions worldwide and a reduction in fossil fuel consumption. The reduction and substitution for the use of fossil fuels are costly, which will lead to the redistribution of benefits on a global scale and inevitably lead to intense conflicts of national interests. Moreover, the international climate negotiations are focused on the formation of international rules to protect the global climate, involving countries of the long-term economic, energy and environmental development, relative rights and responsibilities for each and every country and so on. One great example is Carbon trading. Carbon trading is an approach used to control greenhouse gas emissions by providing economic incentives for achieving carbon dioxide (CO2) emissions reductions. The basic principle of carbon trading is a contract that country can purchase a number of carbon dioxide emissions from other countries if it has the excessive emission. This kind of money can be treated as a penalty that punishes countries that don't care about the environment and Earth. This transaction is calculated in tons per ton of carbon dioxide equivalent (tCO2e), so it is called "carbon trading".

Many companies and organizations are trying to reduce global warming and planning certain strategies and tactics for global warming reduction or mitigation, including advocacy of the use of biodiesel, wind energy, solar energy, nuclear energy. Besides, government advocates citizens to use pure electric vehicles, hybrid vehicles, collects carbon Tax, and controls population. By cooperating with the government, many companies popularized their brand to the public. One of the most successful and widely known company is Tesla. Speaking of sustainable development and climate change, who does not immediately think of Tesla? Tesla is only a dozen years old and it's business idea relied on the superiority of electric motor over gasoline engine, which is either fighting or adapting to a changing climate.

REFERENCES

[1] United States Environmental Protection Agency. Climate Change Science. (????).

[2] NASA's Jet Propulsion Laboratory. 2017. Global Climate Change: Vital Signs of the Planet. (2017).

[3] Berkeley Earth Data Website. 2016. Climate Change: Earth Surface Temperature Data. (2016).

A THE BODY OF THE PAPER

A.1 Introduction

A.2 Related Work

A.2.1 EPA: United States Environmental Protection Agency.

A.2.2 NASA: Climate Change and Global Warming.

A.3 Data set

A.4 Main Techniques Applied

A.4.1 Preprocessing & Data Cleaning.

A.4.2 Data Warehouse.

A.4.3 Classification & Clustering.

A.5 Tools

A.6 Results

A.6.1 Global Temperature Analysis.

A.6.2 Seasonality.

A.6.3 High Trends & Low Trends.

A.6.4 Regional Analysis.

A.6.5 Relationship between Climate Trend and Latitude.

A.7 Applications

A.8 Reference